PLoS one

# Robust Time Estimation Reconciles Views of the Antiquity of Placental Mammals

Yasuhiro Kitazoe[1]*, Hirohisa Kishino[2], Peter J. Waddell[2,3], Noriaki Nakajima[1], Takahisa Okabayashi[1], Teruaki Watabe[1], Yoshiyasu Okuhara[1]

1 Center of Medical Information Science, Kochi Medical School, Nankoku, Kochi, Japan, 2 Graduate School of Agricultural and Life Sciences, University of Tokyo, Tokyo, Japan, 3 South Carolina Cancer Center (SCCC), University of South Carolina, Columbia, South Carolina, United States of America

*Background.* Molecular studies have reported divergence times of modern placental orders long before the Cretaceous–Tertiary boundary and far older than paleontological data. However, this discrepancy may not be real, but rather appear because of the violation of implicit assumptions in the estimation procedures, such as non-gradual change of evolutionary rate and failure to correct for convergent evolution. *Methodology/Principal Findings.* New procedures for divergence-time estimation robust to abrupt changes in the rate of molecular evolution are described. We used a variant of the multidimensional vector space (MVS) procedure to take account of possible convergent evolution. Numerical simulations of abrupt rate change and convergent evolution showed good performance of the new procedures in contrast to current methods. Application to complete mitochondrial genomes identified marked rate accelerations and decelerations, which are not obtained with current methods. The root of placental mammals is estimated to be ~18 million years more recent than when assuming a log Brownian motion model. Correcting the pairwise distances for convergent evolution using MVS lowers the age of the root about another 20 million years compared to using standard maximum likelihood tree branch lengths. These two procedures combined revise the root time of placental mammals from around 122 million years ago to close to 84 million years ago. As a result, the estimated distribution of molecular divergence times is broadly consistent with quantitative analysis of the North American fossil record and traditional morphological views. *Conclusions/Significance.* By including the dual effects of abrupt rate change and directly accounting for convergent evolution at the molecular level, these estimates provide congruence between the molecular results, paleontological analyses and morphological expectations. The programs developed here are provided along with sample data that reproduce the results of this study and are especially applicable studies using genome-scale sequence lengths.

## INTRODUCTION

Despite great progress over the past decade, the evolutionary history of placental mammals remains controversial. While a consensus is emerging on the topology of the evolutionary tree [1–5], although with occasional disagreement [6–8], divergence times remain uncertain. The age of earlier nodes and in particular the root, remain especially uncertain in the absence of definitive placental fossils deeper into the Cretaceous [4–6,9]. Both paleontological and morphological studies suggest that the radiation of placental orders and super orders occurred close to the Cretaceous–Tertiary (K–T) boundary about 65 million years ago (mya) [9,10].

In contrast, molecular studies have suggested markedly older origins for many superordinal groups and that some extant orders diversified before the K–T boundary. The root of living placental mammals has been reported to be in the range of 100–140 mya [4,11–15] even with application of rate-adjustment techniques [4,15]. The molecular consensus of an old root is becoming strong enough that it may become dogma without a close examination of the assumptions being made in such analyses. The true age of the orders and superorders has important implications for determining the overall paleoecology and biogeography of placental mammals during a period that included the breakup of continents and the extinction of terrestrial dinosaurs. Reconciling the different paleontological and molecular divergence time estimates has important implications for basic methodologies that are central to evolutionary study.

The strength of molecular divergence time studies is their potential to draw information from very long aligned sequences of many species. It is widely assumed that a huge amount of sequence data and the approximate rate constancy of sequence evolution

[16] make molecular estimates more reliable than those based solely on fossil data. Such analyses on a genomic scale are generally anticipated to be decisive because they are expected to be free of stochastic noise [17,18].

However, molecular studies acknowledge the problem of misspecification of the model of sequence change, which may result in seriously biased estimation. The relationships among placental orders do vary according to the data used and the taxa sampled [1–8,13–15,19–22]. Most methods of phylogenetic inference rely strongly on probabilistic models of sequence evolution, and neither directly detect nor correct for convergent evolution [23]. When left uncorrected, such "homoplasy" may attract lineages and underestimate certain pairwise distances. This in turn distorts branch (edge) length estimates, which are of primary importance for divergence time estimation. While some methods will detect

. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .

potential convergent evolution, for example Hadamard conjugations and SplitsTree [24], Kitazoe et al. have developed multidimensional vector space (MVS) representation methods to both detect convergent evolution among estimated distances and to also correct for this bias [25–27].

A further problem facing molecular dating is the evolutionary rate constancy, or lack of constancy, over long periods. Since its proposal in 1965 [28], the molecular clock hypothesis has been one of the most hotly debated subjects in evolutionary biology. It is now widely accepted that a molecular clock tends to hold well for closely related organisms, but breaks down with increasing evolutionary divergence. To adjust for the inevitable fluctuation of evolutionary rates, nonparametric [17,18], local clock [29,30], and hierarchical Bayesian [31] methods have been developed. These methods are robust against stochastic fluctuations of evolutionary rates [31–33]. However, they may cause serious problems when pronounced transient changes of rate occur, because they overly smooth such rate changes. In this article, we show the magnitude of such problems using clear worked examples.

To improve the detection of, and robustness to, abrupt rate changes, we have developed a new procedure that minimizes the local variability of the inverse of the evolutionary rate. Just as the effective size of fluctuating populations is represented by the harmonic mean over time, the mean evolutionary rate among lineages is expressed better by the harmonic mean. This approach is especially useful when branch lengths measured in the expected number of substitutions per site (the products of rates and times) are estimated accurately, and there are either rapid transient changes of rate (hence large rate heterogeneity) or a general bias towards a speed up or slow down in rates through time.

Using this new procedure, an analysis of 69 mitochondrial protein sequences (3660 amino acid sites in total) from placental mammals identified a rapid acceleration of evolutionary rate for the lineage directly leading to the common ancestor of Supraprimates and an even more marked one for the lineage leading to Laurasiatheria. This acceleration was followed closely by a strong deceleration, which persisted in nearly all lineages of Laurasiatheria. In contrast, almost all lineages of Afrotheria and Xenarthra seem to have retained rates similar to that of the root. This view is in marked contrast to current rate-change penalty functions. The robustness of the new procedure is assessed using simulations that show the types of change that most concern biologists; speedups or slowdowns through time, transient rate changes, and rate changes that are do not follow a normal or transformed normal distribution, as well as stochastic variation. A revised estimate of the origin of placental mammals is as young as 84 mya, which is much more recent than current estimates using molecular data. The inferred age of deeper splits in the placental tree are compared with the rate of occurrence of new species from the North American fossil record. These two sources of data are far more congruent than is suggested by using current, possibly strongly misleading, dating methods.

## RESULTS

### Revised rates and times of mitochondrial protein sequences from placental mammals

Mitochondrial protein sequences are used widely in phylogenetic studies and have been particularly popular in studying placental mammals. A desirable feature of these data is relatively long sequences, good taxon sampling and very little missing data. Following alignment, we retained 3660 amino acid sites present in all of 62 placental mammals plus seven outgroup taxa (Table S1 gives the accession numbers of protein sequences). In contrast,

large nuclear protein datasets of mammals with diverse taxon sampling [e.g., 3, 4] show a large proportion of missing data. Such missing heterogeneous data may lead to complex systematic errors in distance estimation [34]. This is particularly relevant since MVS analyses work best with low stochastic noise (hence relatively long sequences), diverse taxon sampling with a considerable proportion of taxa showing minimal convergent evolution, and minimal bias of the input distances from sources such as missing data.

We adopted a standard two-step procedure to estimate divergence times. The first step is to estimate the phylogenetic tree with unconstrained branch lengths in units of expected numbers of substitutions per site. Given the problems with convergent evolution in mitochondrial data [4,5,22,25], we used the MVS procedure to correct the input distances for convergent evolution. A variant of the core set approach [27] to MVS was used in this instance (Materials and Methods). Fixing the topology to the tree obtained by MVS, which is very similar to trees obtained by previous authors using varied data sets [2–5], branch lengths were reestimated using maximum likelihood (ML) in the program PAML [35]. We used the JTT [36]+Gamma model with $\alpha = 0.5$ for these analyses. The second step uses the tree and its branch lengths, some fossil constrained nodes, and a penalty or cost function to dampen rate changes, to infer divergence times of all nodes and evolutionary rates along all branches.

Three cost functions, $F_{ADD}$, $F_{LOG}$, and $F_{IR}$, were applied to the MVS and ML trees. These functions penalize the fluctuation of rates, and do this on either a linear scale (here called the ADD function), on the log rate (LOG), or on the inverse rates (IR), respectively (see equations (1), (2), and (4) in Materials and Methods). The tree estimated by MVS and $F_{ADD}$ is denoted MVS-$F_{ADD}$ and so on (while ML-$F_{ADD}$ indicates use of ML branch lengths).

To calibrate these trees, we used eight fossil constraints, all taken from previous studies (see Figure 1 caption) [4,15]. The divergence times were estimated by then minimizing each cost function subject to these constraints (see Materials and Methods). The CI was calculated using a likelihood interpretation (see Materials and Methods) of the cost function residual analogous to the method used in Multidivtime [31]. This captures the error caused by deviations of the tree's branch lengths away from their expected values and includes the unpredictability of rate changes, which might mimic Brownian motion, for example. It will also incorporate variability arising from ancestral polymorphism. Polymorphism will cause fluctuations of the edge lengths of the tree when the analysis has multiple nodes constrained by fossil or other data [c.f. 13]. Generally, the time at the most recent common ancestor of sequences from different species is older than the time at speciation. The extent of this difference varies among internal nodes due to both the stochastic nature of the coalescent and factors governing its expected magnitude, such as population size. The cost function does not take account of this bias, but the CI does include the variance arising from ancestral polymorphism at internal nodes.

The MVS-$F_{IR}$ analysis (Figure 1A) estimated the age of the root at 84.2 mya with a 95% confidence interval (CI) of 80.7–88.4 mya (Table 1). This time is much more recent than the result of the ML-$F_L$ analysis (Figure 1B), i.e., 122.2 (95% CI, 112.2–144.7) mya. Note that, ML-$F_L$ is giving results consistent with that reported in earlier studies [1–5,12–15], indicating that the younger root for placentals returned using the function $F_{IR}$ (see below) is due to the method and not the data. The difference between the $F_{LOG}$ and the $F'_{LOG}$ function in table 1 is explained in Materials and Methods. It is the $F_{LOG}$ method that most closely approximates the Brownian motion assumed by Multidivtime.
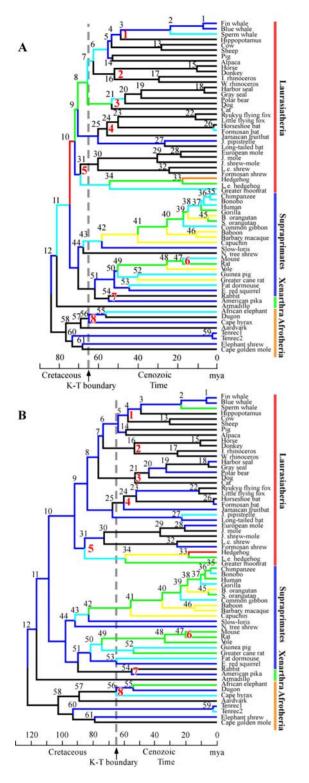
**Figure 1. Molecular divergence times and evolutionary rates of placental mammals.** A and B show the MVS-$F_{IR}$ and ML-$F_{LOG}$ trees, respectively. The numbers 1–61 denote the ancestral nodes. The red numbers **1–8** indicate the internal nodes with fossil constraints which are as follows: **1**, 49–61; **2**, 52–58; **3**, 45–63; **4**, 43–60; **5**, <63; **6**, >12; **7**, 36–55; **8**, 54–65 (all in mya) [see 4, 15]. The colors of the branches denote inferred evolutionary rates (in units of $\times10^{-9}$/site per year) as follows: black, <0.2; dark blue, 0.2–0.3; light blue, 0.3–0.4; green, 0.4–0.5; brown, 0.5–0.6; yellow, 0.6–0.7; and red, >0.7.
doi:10.1371/journal.pone.0000384.g001

Figure 1 shows the estimated ancestral rates across the tree, while Figure 2 represents the inferred evolutionary rates going from the root to a terminal. The single instance of root to fin whale is shown in figure 2, while figure S1 traces the evolutionary rates inferred by different methods along seven representative lineages. The MVS-$F_{IR}$ tree identified an abrupt acceleration of evolutionary rate near the common ancestor of both Supraprimates and Laurasiatheria, then a very strong acceleration in just the ancestral lineage of Laurasiatheria (Figures 1A and 2A). This acceleration was followed closely by a strong deceleration along nearly all lineages of Laurasiatheria. In contrast, almost all lineages of Afrotheria and Xenarthra have retained rates close to that of the root. There are also sporadic later accelerations, for example that among hedgehogs and Moon rat (Figure 1A).

The cost function $F_{IR}$ detected an acceleration–deceleration pattern near the base of Laurasiatheria using both the MVS and ML branch lengths (the red lines of Figure 2), whereas the function $F_{LOG}$ showed a far more flat prediction of generally lower evolutionary rates that led to older root times (the green lines of Figure 2). Even more extreme, the cost function $F_{ADD}$ inferred gradually decreasing rates in all deep branches of the MVS and ML trees (the blue lines of Figure 2), and the root time tended to infinity. Table 1 summarizes the root time values estimated by various models (all the node divergence times from these models are listed in Table S2).

Table 1 also compares the inferred times with those obtained from *r8s*, a popular program for estimating absolute rates [17, 18]. This program has two options of for its cost functions, $F_{ADD}$ and $F'_{LOG}$ (equation (3) in Materials and Methods). It also takes account of stochastic noise caused by the finite length of the sequences used to derive the branch lengths using a penalized-likelihood approach. The fitting to the estimated branch lengths is expressed by the log likelihood, while the weight for the cost function is estimated using cross validation. In our mitochondrial dataset, with its long sequences, the root time values changed little due to the cross validation effect (the reason is explained in Materials and Methods). Further, root divergence time estimates returned by *r8s* do not show confidence intervals in table 1, since *r8s* only assesses the sampling variance of node times due to stochastic errors in branch lengths due to finite sequence length (it does this via the bootstrap) and does not include the effect of stochastic variation of evolutionary rate. This source of error remains as sequence lengths go to infinity, and will be dominant with long sequences.

## Agreement with fossil data

Figure 3 represents the chronological distribution of the internal node density for the MVS-$F_{IR}$ tree (Figure 1A) compared with the rate of new species appearing in the well-studied North American fossil record. To avoid potential bias of divergence times in the molecular tree due to species sampling, the scaling constant was determined by a least-squares fit to the fossil data in the period 85–50 mya.

The MVS-$F_{IR}$ analysis reconciles molecular with fossil data in two ways. First, the chronological distribution of the internal node density is clearly largely consistent with the rate of appearance of novel fossil species. This is despite the fact that the paleontological data assessed are limited to the well-studied North American record [10]. Both the MVS-$F_{IR}$ tree and the fossil record suggest accelerating taxonomic diversity near the K–T boundary, rather than a longer slower buildup [2–5,12–15]. Use of the less robust $F_{LOG}$ criterion seems to suggest a prolonged increase in diversity that agrees far less with the quantitative fossil record. This congruence with fossils does not automatically show that the

**Table 1.** The age of the root of placental mammals

| Tree model | Cost function | | | | | |
|---|---|---|---|---|---|---|
| | $F_{IR}$ | $F_{ADD}$ | $F_{LOG}$ | $F'_{LOG}$ | $F_{ADD}$(r8s) | $F'_{LOG}$(r8s) |
| MVS | 84.2(80.7, 88.4) | Infinity | 105.0(97.0,117.0) | 91.8(84.5,104.7) | 160.5 | 91.1 |
| ML | 106.4(102.4,110.6) | Infinity | 122.2(112.2,144.7) | 112.0(102.0,121.2) | 122.2 | 111.6 |

The age of the root using MVS and ML branch lengths was estimated after minimizing various penalty functions with the same fossil constraints. All times are in mya, and 95% confidence intervals were estimated by the sum-of-squares method described in Materials and Methods. The times estimated by the *r8s* program use cross-validation and penalized likelihood methods with essentially the same penalty functions [18], but do not report comprehensive errors.
doi:10.1371/journal.pone.0000384.t001

combination of MVS and $F_{IR}$ is the best way to analyze this data, but it does reframe the discussion of the relationship between the fossil and molecular times into one where the molecular dates are being looked at far more critically.

Second, and more indicatively, the cost function $F_{IR}$ resolves incongruence among the fossil constraints/inferred divergence times in different parts of the molecular tree. The best fossil constraints in Laurasiatheria suggest much older times than constraints in other parts of the tree [4,37]. The MVS-$F_{IR}$, MVS-$F_{LOG}$, and MVS-$F_{ADD}$ trees with all constraints give the root as 84.2, 105.0, and ∞ mya, respectively (Table 1). Removing the whale–hippo constraint gives 82.4, ∞, and 91.2 mya; removing the horse–rhino constraint gives, 82.8, 105.0, and ∞ mya; and removing both sets of constraints gave 82.9, 87.2, and ∞ mya.
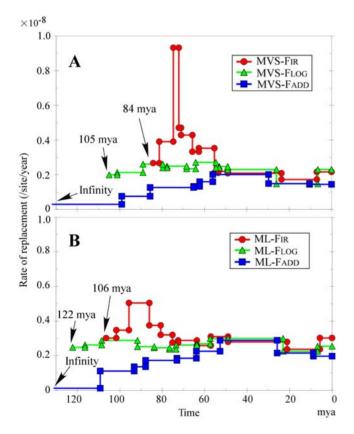


**Figure 2. Evolutionary rate changes from the root to the fin whale.** Figures A and B, respectively, trace the estimated rates along edges on analyses using MVS and ML branch lengths. The root time of of the $F_{ADD}$ analysis in Figure A and B was set at a large value (400 mya) because the numerical calculation continues towards an infinite root time.
doi:10.1371/journal.pone.0000384.g002

Thus, in this case only the $F_{IR}$ tree is insensitive to "constraint sampling".

Finally, there is a good fossil calibration for tarsier [4] that was withheld because it is not used in both references 4 and 15. It suggests that human and tarsier split around 50–60 mya and the molecular trees suggest the human-loris split was not much earlier (probably <5 myr earlier). Despite our use of whale–hippo and horse–rhino calibrations, the MVS-$F_{IR}$ tree (Figure 1) gives a human-loris split close to this age, in contrast to earlier studies [4,15,33].

## Sensitivity and robustness of $F_{IR}$ assessed by simulations

We highlight two distinct properties of the new function $F_{IR}$ with the help of evolutionary simulations and worked examples. The first property is its ability to detect a transient acceleration of evolutionary rate. Such an effect might be caused by a burst of positive selection and/or a bottleneck in population size. The second is to assess the effects of both stochastic fluctuations and systematic bias on the robustness of estimated times. Here, we model bias in the form of either a general slowdown or a general acceleration of evolutionary rate across the whole tree.
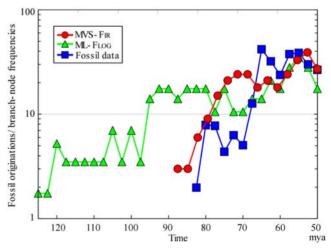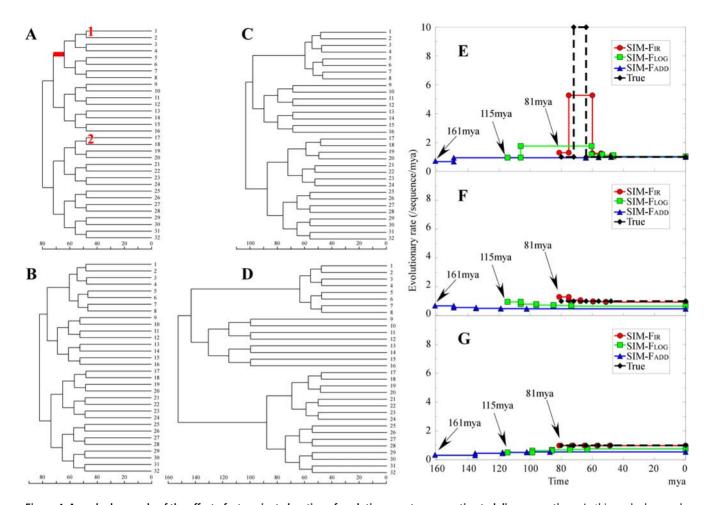


**Figure 3. Comparing molecular and fossil records of placental diversification.** The blue squares show the rate of the appearance of new species based on the fossil record [adapted from 3]. At present, such quantitative data are limited to the well-studied North American record. The red circles represent the chronological distribution of node density (or splits) on the MVS-$F_{IR}$ tree (Figure 1A). The green triangles represent split frequency on the ML-$F_{LOG}$ tree (Figure 1B). The node density is given by the number of nodes in an 8 myr sliding window. A scaling constant for the molecular frequencies was calculated via a least-squares fit to the fossil data in the period 85–50 mya.
doi:10.1371/journal.pone.0000384.g003

**Figure 4. A worked example of the effect of a transient elevation of evolutionary rate upon estimated divergence times.** In this worked example using a symmetric 32-taxon tree (Figure A), a global molecular clock holds, except for a short-term increase in evolutionary rate along one branch (the red line in Figure A). The true root time was set to 80 mya, and the times at the internal nodes are 48, 56, 64, and 72 mya. The deep internal branch (the red line in Figure A) is given an evolutionary rate ten times that of the remaining edges. The various cost functions were minimized subject to two calibrated nodes (the red numbers **1** and **2** in Figure A), using the exact branch lengths of this example as input data. The cost functions $F_{IR}$, $F_{LOG}$, and $F_{ADD}$ inferred the weighted trees of Figures B, C and D, respectively. Figures E–G show the trace of evolutionary rates along the lineages from the root to taxa numbers 5, 9, and 25 of Figure A, respectively. The inferred age of the root for each cost function is shown with an arrow. The function $F_{IR}$ recovered the original pattern of rate change, whereas the other two functions inferred far more gradual changes, which resulted in a substantial overestimation of the root time.
doi:10.1371/journal.pone.0000384.g004

We first modeled a strong instantaneous acceleration as an analogue to what is inferred by $F_{IR}$ to have occurred ancestral branch leading to Laurasiatheria. We simulated a 32-taxon symmetric tree in which a molecular clock holds except for an abrupt elevation (by a factor of 10) of evolutionary rate along a short internal branch (the red line in Figure 4A). This example is useful for demonstrating the efficacy of the $F_{IR}$ function and something similar appears on both the MVS-$F_{IR}$ and ML-$F_{IR}$ mitochondrial trees. The times at the internal nodes were set to 48, 56, 64, and 72 mya, and the root time was set to 80 mya. On this weighted tree, the cost functions were minimized with two constrained node times, which corresponded to two fossil calibration points. Only the $F_{IR}$ function accurately estimated the true divergence times and appeared robust against the abrupt change (Figure 4B). In contrast, the functions $F_{LOG}$ and $F_{ADD}$ inferred gradual rate changes (Figure 4E), which are erroneous and lead to overestimation of the root time along with that of many other nodes (Figures 4C and 4D). Table 2 summarizes the root time values inferred by various models. Table 2 includes the result of *r8s* with the cross-validation method.

We next simulated stochastic rate fluctuations by themselves, plus either a prevailing slowing down or acceleration of rates through time. Such simulations are distinct from a Brownian-type process, and are used to gauge the general robustness of the functions. To impose rate fluctuations on the same basic tree as

**Table 2.** Estimated age of the root in the presence of an abrupt acceleration then deceleration of ancestral rates

| True value | Cost function | | | | | |
|---|---|---|---|---|---|---|
| | $F_{IR}$ | $F_{ADD}$ | $F_{LOG}$ | $F'_{LOG}$ | $F_{ADD}$(r8s) | $F'_{LOG}$(r8s) |
| 80 | 81.3 | 161.1 | 114.5 | 101.7 | 160.5 | 108.7 |

These estimates are from a worked example with an abrupt acceleration then deceleration (the red line of Figure 4A) which occurs deeper in the tree than the calibration points. The true root age is 80.0 mya. The estimated values are made by minimizing the cost functions (that also appear in table 1), but upon the weighted constrained tree of figure 4A.
doi:10.1371/journal.pone.0000384.t002

Figure 4A, but without the abrupt rate change, the age of the root was set to 100 mya. We assumed infinite length of sequences and that the branch lengths are known without uncertainty. Next a branch was randomly selected proportional to its duration in time and all its descendant branch lengths were multiplied by a factor (chosen randomly) from a uniform distribution of range 0.5–1.5. A total of 25 such rate changes were placed on the tree to give the final branch lengths for that tree. These branch lengths were used as the input data and the various cost functions were minimized with on calibrated internal node, and then determined all node times. We repeated this procedure 600 times to obtain the mean and standard error of the root time. We also simulated the two other cases. The only difference was the range of the uniform distribution used. Using a range of 1.0–1.75 the model is biased toward rate acceleration through time, while using and 0.25–1 induces a rate slow down through time (a probable situation in the placentals generally). As table 3 shows, the function $F_{IR}$ gave reasonable estimates with a bias towards deceleration, whereas $F_{ADD}$ gave an infinite root time in 139 of 600 samples. These undefined root-time values were set arbitrarily to 200 mya before calculating the mean and standard error.

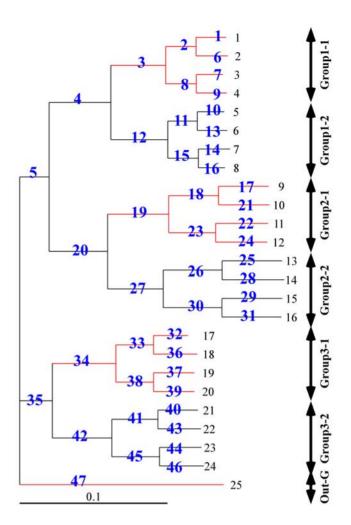## Reduced bias in branch length estimation by the MVS model

The MVS model was shown previously to recover the correct tree in a simulation with two strongly convergent lineages [26] that were grouped erroneously by standard methods (including the neighbor joining (NJ) [38] and ML [35] methods). Here, we examined a different question; this is how well the MVS method recovers the true branch lengths, which are of primary importance when estimating divergence times. We simulated the evolution of a sequence of $10^4$ amino acid sites, following the tree depicted in Figure 5. The model of amino acid substitutions used was the JTT [36]. Convergent evolution was then imposed on this data. This was done by sharing parts of the sequences among three ingroup clades plus the outgroup. The red lines in Figure 5 indicate which
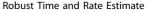


**Figure 5. A simulation of convergent evolution.** Amino acid substitutions were evolved on the shown weighted tree using the JTT model [36] of amino acid substitutions. After this, 30% of sites were swapped between the four lineages indicated by the red lines.
doi:10.1371/journal.pone.0000384.g005

lineages shared sequence and were therefore subject to a form of convergent evolution.

Pairwise distances were then estimated from the terminal sequences obtained above using the same JTT model. A modified MVS core-set procedure [27] was able to recognize that the tree consisted of three groups of eight sequences which had additive distances within each group and the outgroup. The MVS procedure then converted the distances within the three ingroups into three sets of perfectly additive distances and sequentially combined them to form a single core set (Materials and Methods). The final MVS tree was obtained by modifying the distances between this single ingroup core set and the outgroup following the rules described in Materials and Methods.

The branch lengths recovered by the MVS model reproduced the true values accurately (Figure 6). The distribution around the diagonal line in Figure 6 represents the stochastic fluctuation of the estimated distances; that is, the magnitude of this fluctuation did not change after the simulation was rerun without convergent evolution. In contrast, the NJ and ML trees returned branch lengths that were affected clearly by convergent evolution. The worst affected branch lengths were ancestral to the groups showing convergent evolution (underestimated) or else were leading to the

**Table 3.** The age of the root inferred by functions $F_{IR}$, $F_{LOG}$, and $F_{ADD}$ on simulated data with random auto-correlated rate changes

| Rate range | Cost function | | |
| --- | --- | --- | --- |
| | $F_{IR}$ | $F_{LOG}$ | $F_{ADD}$ |
| 0.25–1.0 | 111.3±11.1 | 125.1±12.7 | 172.0±33.9 |
| 0.5–1.5 | 100.7±4.4 | 103.3±3.9 | 103.6±5.1 |
| 1.0–1.75 | 93.2±2.0 | 94.0±2.1 | 94.3±2.4 |

In these simulations, the length of a branch (on what tree!) is selected randomly proportional to its true duration in time. A rate adjustment (change) factor is then chosen randomly form a uniform distribution and all its descendant branch lengths were then multiplied by this factor. A total of 25 such random rate changes were placed on the tree, then the branch lengths (measured in the product of rate and time) of the weighted tree were passed on to the time estimation procedures. The whole procedure was repeated 600 times to obtain the average root time and standard error. Three different ranges were used for the uniform distribution of rate changes. The first has a range of 0.25 to 1 and represent a strong persistent bias towards rate deceleration. A range of 0.5 to 1.5 gives minimal rate change bias, but retains stochastic fluctuations. Finally a strong acceleration effect is achieved by the use of the range 1.0–1.75. The function $F_{ADD}$ gave an age of the root tending to infinity on 139 of the weighted trees simulated under the deceleration model. In such cases the root time is set to 200 mya in order to allow the mean and standard deviation of this cost function to be calculated.
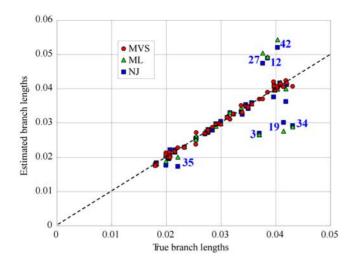doi:10.1371/journal.pone.0000384.t003

**Figure 6. Branch lengths estimated by MVS, NJ, and ML for the example of convergent evolution in figure 5.** The tree topology inferred by these methods was identical to the tree that generated the data, so the estimated branch lengths are plotted against their true values. The blue numbers show the branch index, as used on Figure 5, of outliers. Note, all the outliers are internal branches ancestral to the four lineages undergoing convergent evolution or are ancestral to the sister group to these lineages. Branch lengths ancestral to the groups undergoing convergent evolution are underestimated by the ML and NJ methods, whereas those ancestral to their sister taxa are over-estimated.
doi:10.1371/journal.pone.0000384.g006

sister groups of the groups affected by convergent evolution (overestimated).

## DISCUSSION

We begin the discussion by examining why the current cost functions, $F_{LOG}$ and $F_{ADD}$, overestimated the age of the root in the worked example with strong rate heterogeneity in the form of a short term highly elevated evolutionary rate. It is also important to examine the profile of the cost function around their minimal values for the age of the root. The functions $F_{LOG}$ and $F_{ADD}$ showed asymmetric behavior around the estimated root time, even in worked examples with a perfect molecular clock (data not shown), whereas the $F_{IR}$ profile was symmetric and parabolic in shape. The asymmetry seen with $F_{LOG}$ and $F_{ADD}$ increased in response to a general bias towards deceleration of rates through time (Figure 7A). In particular, the function $F_{ADD}$ showed a monotonic decrease with respect to increasing age of the root, that is, its estimate tended to infinity. Thus, asymmetric behavior of a cost function seems to be a symptom of unstable estimation of the age of the root. A similar strong asymmetry appeared in the profile of the current cost function with respect to the age of the root of the mitochondrial tree of placental mammals (Figure 7B), irrespective of whether MVS or ML branch lengths were used. Comparing Figures 7A and 7B suggests that bias due to decelerating evolutionary rates occurred in the evolutionary history of placental mammals, and is impacting the ability of current methods to estimate the correct age of the root.

Other approaches to divergence time estimation are under active development. For example, Drummond et al. [39] recently developed a set of Bayesian procedures to jointly estimate divergence times, evolutionary rates, and the tree topology. The uncorrelated variable rate model they use, which assumes independence of evolutionary rates between branches, may also be able to identify instantaneous accelerations. However, the
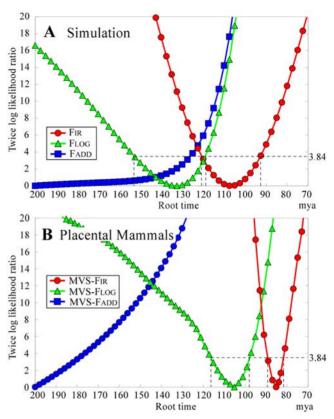


**Figure 7. Log-likelihood profiles of the $F_{IR}$, $F_{LOG}$, and $F_{ADD}$ cost functions about the estimated ages of the root.** This figure shows the profile of the cost function with respect to the age of the root estimated by three different cost functions. Because the root age estimated by function $F_{ADD}$ was going to infinity, it was set to 200 mya for illustrative purposes. Figure A is a single example from a tree simulated under the scenario of random auto-correlated changes of rate moving towards the tips, strongly biased towards a deceleration of evolutionary rates as time progresses (from the set of simulations used for Table 3). The true age of the root was 100 mya. Figure B shows the results using the MVS tree derived from the mitochondrial sequences of placental mammals. The dotted line indicates the 95% confidence interval of the estimated age of the root using the sum-of-squares approach described in Materials and Methods.
doi:10.1371/journal.pone.0000384.g007

magnitude of rate acceleration may be underestimated because hierarchical Bayes estimates can generally be regarded as shrinkage estimators [40]. An important future direction is to assess the performance of the various estimators in simulations where rate variability is a mixture of an autocorrelated process plus rare, but strong, instantaneous accelerations.

Our refined approaches resolve apparent contradictions between the quantitative molecular and paleontological data of placental mammals. Given such agreement, there is no need for ancillary hypotheses such as the long fuse model [2,5,9,12–15,41] to explain the lack of any positively identified fossils of modern placental mammals prior to about 75 million years ago. In addition, dates of markedly less than 100 mya for the root of placental mammals also bring into question earlier hypotheses that traditional continental drift models explain the geographic distribution of the four major groups of placental mammals [2–5,13,42]. We anticipate that the $F_{IR}$ cost function developed in this paper will provide an improved methodology for a wide range of molecular studies because its robustness to rapid fluctuations of rate is essential to understanding events such as adaptive evolution.

For example, because acquisition of new molecular functions can be achieved in a few million years after gene duplications [40], the duration of an inflated evolutionary rate may be surprisingly short, and, correspondingly, difficult to detect. We hope our new methods will be beneficial in such situations and will lead to a richer understanding of molecular evolution.

## MATERIALS AND METHODS

### Autocorrelated variation of inverse rate

Because the branch lengths of a phylogenetic tree are estimated from the data as the product of the evolutionary rate of a branch and its time duration, these two factors cannot be directly estimated separately. It has become standard to use loose constraints to accommodate uncertainty and it is often wise to exclude constraints if there is no firm basis for them. Information on the age of some internal nodes may be fairly directly available from the fossil record. An example of this is the horse-rhino split [13]. If the molecular clock [16] governs the process of molecular evolution well, we can accurately estimate the times at other nodes given just a single reliable calibration point. However, the assumption of the molecular clock is often rejected by the fit of a clock-like tree to the data.

The most widely used assumption in order to model the evolutionary rate changes away form being constant (or away from a clock) is to introduce a stochastic process. For example, Sanderson [17] suggested the following cost function be minimized:

$$F_{ADD} = \Sigma_n (R_n - R_{a(n)})^2 \qquad (1)$$

with rates $R_n$ and $R_{a(n)}$ being from the $n$th branch and its ancestral branch $a(n)$, respectively. The exponent of the negative cost function $\Pi_n \{(2\pi)^{1/2}\sigma\}^{-1} \exp\{-(R_n - R_{a(n)})^2/2\sigma^2\}$ implies that this function can be interpreted assumes *a priori* that rate changes follow a random walk with independent increments randomly drawn from a normal distribution. Likewise, Thorne et al. [31] considered a cost function of the type

$$F_{LOG} = 2\Sigma_n (\log R_n - \log R_{a(n)})^2 / (T_n + T_{a(n)}) \qquad (2)$$

in the framework of the hierarchical Bayes procedure. Here $T_n$ is the time interval of the $n$th branch (the current version of Multidivtime [31] uses rates at the nodes, instead of average rates along branches, as the free parameters). This can be interpreted as an assumption that the log rate undergoes Brownian motion. That is, independent increments of normal random variables, whose variances are proportional to the average time duration between the pair of branches (that is, the average length in time of the two branches). Recently, Sanderson implemented the unweighted cost function

$$F'_{LOG} = 2\Sigma_n (\log R_n - \log R_{a(n)})^2 \qquad (3)$$

into his program *r8s* [18].

However, the estimated rates based on any of the above cost functions may overly smooth the change of evolutionary rates when a pronounced transient change of rate has occurred. In turn, biased estimates of evolutionary rates will lead to biased estimates of divergence times. Here we propose a new cost function, which penalizes the local rate deviation from the harmonic mean. For simplicity, we ignore the stochastic variance of branch length [17,18,31], which goes to zero with increasingly long sequences.

Denote the branch length and inverse rate of the $n$th branch by $B_n$ and $Q_n$ $(= T_n/B_n)$, respectively. The local rate variability ($F_{IR,n}$) of two successive branches can be expressed as the variance of $Q_n$ and $Q_{a(n)}$ around their average value, $A_n = (Q_n B_n + Q_{a(n)} B_{a(n)})/(B_n + B_{a(n)})$, which is equal to the inverse of the harmonic mean of $R_n$ $(= B_n/T_n)$ and $R_{a(n)}$, with the weights $B_n$ and $B_{a(n)}$. That is, $F_{IR,n} = \{(Q_{IR,n} = \{(Q_n - A_n)^2 B_n + (Q_{a(n)} - A_n)^2 B_{a(n)}\}/(B_n + B_{a(n)})$. Because $F_{IR,n}$ is re-written as $F_{IR,n} = (Q_n - Q_{a(n)})^2 W_n$ with the weight $W_n = B_n B_{a(n)}/(B_n + B_{a(n)})^2$, the inverse-rate cost function, $F_{IR}$, is defined as the weighted average of $(Q_n - Q_{a(n)})^2$ over all ancestor–descendant branch pairs in the tree:

$$F_{IR} = \Sigma_n F_{IR,n} / \Sigma_n W_n = \Sigma_n (Q_n - Q_{a(n)})^2 W_n / \Sigma_n W_n, \qquad (4)$$

which contrasts with previous expressions in terms of $R_n$.

Because $F_{IR}$ places a smaller penalty on abrupt rate changes than do previous models, it can confine an abrupt change close to where it occurred on the tree. In contrast, $F_{ADD}$ and $F_{LOG}$ *a priori* put stress upon the smoothness of evolutionary rate change from one branch to the next, and this may propagate the effect of an abrupt rate change over successive branches. In this article, the variable $(T_n)$ (the divergence times) were estimated by minimizing the cost functions (equations 1–4) subject to the (fossil) constraints without introducing other modifiers such as cross validation.

At first glance, it seems most intuitive to use a penalty such as $(rate_1 - rate_2)^2$; this has been the implicit assumption until now [31] and is the basis of published programs such as *r8s* and *Multidivtime* [17,18,31]. It rests on the implicit expectation that any reasonable type of smoothing of changes in rate will give a similar answer. However, a simple linear form for the difference in rates is misleading, and this is illustrated by considering the general problem of estimating velocity (rate) given distance (in our case branch length). Going from point $a$ to $b$ at velocity $v_1$, then back at velocity $v_2$, the average velocity $(v_{av})$ is equal not to the standard arithmetic mean, but to the harmonic mean, $1/v_{av} = (1/v_1 + 1/v_2)/2$. Further, going from point (node) $a$ to $b$ at velocity (rate) $v_1$, and from point $b$ to $c$ at velocity $v_2$, then the distances traveled ($d_{ab}$ and $d_{bc}$) need no longer be equal so we must use weights. Specifically, $v_{av} = (d_{ab} + d_{bc})/(t_1 + t_2)$, where $t_1 = d_{ab}/v_1$ and $t_2 = d_{bc}/v_2$. As a result, $1/v_{av} = (d_{ab}/v_1 + d_{bc}/v_2)/(d_{ab} + d_{bc})$. This is the same form as the average quantity $A_n$ used to obtain $F_{IR}$ (equation 4). The use of the harmonic mean becomes important when $v_1$ differs greatly from $v_2$.

### The penalty in hierarchical Bayes and penalized likelihood approaches

The Bayesian approach estimates the divergence times and evolutionary rates in the form of a posterior distribution, which is summarized approximately as $P(\boldsymbol{B}|\boldsymbol{R},\boldsymbol{T}) \exp(-\lambda F)$. Here $\boldsymbol{B}$, R, and $\boldsymbol{T}$ are the vectors of the estimated branch lengths, evolutionary rates, and divergence times, respectively, while $F$ is the cost function. The value of $\lambda$ expresses the weight for the penalty, and is called a hyperparameter. Introducing the distribution for the hyperparameter is done via a so-called hyper-prior; the hyperparameter is then estimated concurrently with the posterior distribution. The penalized likelihood approach maximizes log $P(\boldsymbol{B}|\boldsymbol{R},\boldsymbol{T}) - \lambda F$. The weight for the penalty $\lambda$ is estimated using cross-validation. When the sequences are long enough, for example concatenated mitochondrial protein sequences, it is often safe to assume that the branch lengths are estimated accurately, that is with minimal stochastic error (but not necessarily without systematic error, something we address in this paper using MVS). Accordingly, it is assumed that the products of rates and times are known exactly. Thus in a situation of long, or very long sequences

(e.g. genomic alignments), the results returned by using just the penalty function will converge to those returned by using either a Bayesian or a penalized likelihood approach with the same type of cost function.

## MVS core-set correction of biased evolutionary distances

In the MVS method, the additivity of evolutionary distances is converted using the expected orthogonality among branch vectors in a multidimensional Euclidean space [25]. If the estimated pairwise distances do not satisfy additivity, MVS provides an index to measure deviations from orthogonality without specifying a tree structure. This index enables one to diagnose whether a distance matrix is compatible with a tree and to modify the biased distances until they satisfy orthogonality, that is, with a zero value for the index. Here, a revised core-set approach, that is, an extension of [27], is applied to correct the observed pairwise distances for convergent evolution.

The first step of this approach is to divide the set of taxa into subgroups and to correct the distances between pairs in each subgroup. We decompose the taxa based on partitions with both strong biological support and high bootstrap support. The deviation from additivity of pairwise distances is much smaller within each subgroup than across the whole distance matrix. If an anomalously large deviation is observed associated with a single taxon within a subgroup, that taxon is removed temporarily from the analysis. When the only deviations between distances within a subset of taxa are judged to be due to stochastic noise, the pairwise distances are modified by solving the equation of motion (a method using a many body kinetic equation in physics), which uses the index of the deviation from additivity as the potential energy [25,26]. The modified distance matrix within a subgroup is interpreted as the true additive distance matrix (core set) achieved by minimal modification of the original distances. Generally, the modified distances are very close to the distances estimated by the NJ method for just these taxa. Then, the excluded taxa are deposited into this core set tree by solving the equation of motion with the same potential function. Assuming that convergent evolution and long-branch attraction are the main source of deviation from additivity, we allow for only positive corrections of the biased distances between the core set and excluded taxa (that is, distances can only get bigger).

The second step is a clustering of subgroups. The pairwise distances between two core-set groups are corrected by assuming attractions between them (that is, some distances are underestimated) and by minimal enlargement of the distances between the two groups to satisfy additivity. In this way, the core-set approach used in this article is in the direction opposite to that used in an earlier procedure that proceeds from the whole tree down to subtrees [26]. This revised method seems more appropriate here because the branching pattern within subgroups becomes more apparent and there are smaller deviations from additivity.

Conceptually, the first step of MVS core set is to decompose the whole tree into putative monophyletic groups that show minimal bias internally. The approach developed as part of this particular study resembles $k$-means clustering, that is, minimizing the deviation (variance) within groups while maximizing the deviation (variance) between groups. At present, the decomposition is done in a supervised, as opposed to an automatic, manner. In the second step, and when it is difficult to connect two groups definitely, we first exclude taxa with the largest deviations from additivity and then complete a new core set for the remaining sequences. We then locate the excluded sequences on the core set tree using minimal modification of distances (a parsimony-like criterion) under the constraint of only enlarging the distances identified as being biased (thus the only biases are assumed to be due to attraction).

When analyzing the mitochondrial protein sequences, we estimated pairwise distances initially under the JTT+Gamma model with $\alpha = 0.5$. Given moderately high bootstrap support, the placental tree was decomposed into four major groups: Laurasiatheria, Supraprimate, Xenarthra, and Afrotheria. Laurasiatheria was decomposed into five subgroups, namely Cetartiodactyla, Perissodactyla, Carnivora, Chiroptera, and Euliptotyphyla. Supraprimates was decomposed into two subgroups: Primates and Glires. The core-set analysis began by generating additive distances within these subgroups, after which the clustering procedure described above was applied to connect these subgroups. In this analysis, we analyzed sequences of Xenarthra and Afrotheria together in a single core-set. This is because a relatively short internal branch separates them and Xenarthra contains but a single sequence in this analysis. Finally, we tried to modify the distances between placental mammals and the outgroups, but we could not determine the position of the root with confidence because of very strong attractions between many placental mammals and the outgroups. The consensus [3–5,15] has been that the most probable placement of the root separates Afrotheria from the other placental mammals. The most likely alternative placement for the root is between Xenathra plus Afrotheria (Atlantogenata [2,20]) and all other placentals (Boreotheria). Recent analyses including LINE sequences and indels [43] support this view. When we place the root in this position, the divergence times, including that of the root, change little (results not shown).

Further details of the current MVS procedure of distance modification are documented in Methods S1. The derived additive distance matrix was converted to a Newick tree file with the help of the NJ method, and this was used as the input data for divergence time estimation by the cost functions.

## Log-likelihood profiles of the cost function with respect to the age of the root

We note that least-squares estimators can be interpreted as ML estimators when errors in the data follow a multi-normal distribution. Accordingly, the least-squares residual can be treated as twice the log-likelihood ratio for the three cost functions $G_{\mathcal{J}}(T_{\text{root}})$ ($\mathcal{J} = $ IR, ADD, or LOG); that is, as $G_{\mathcal{J}}(T_{\text{root}}) = \mathcal{N} \log\{F_{\mathcal{J}}(T_{\text{root}})/F_{\mathcal{J}}(T_M)\}$. Here, $F_{\mathcal{J}}(T_{\text{root}})$ gives the minimum residual value with all internal node times reestimated, except for the root time $T_{\text{root}}$, and it becomes truly minimal at $T_{\text{root}} = T_M$. Here, $\mathcal{N}$ is the number of terms in the cost function, which is equal to $2t-4$, where $t$ is the number of tips on the tree (here $t = 62$, so $\mathcal{N} = 120$). The 95% CI can be obtained using the threshold of $G_{\mathcal{J}}(T_{\text{root}}) = 3.84$. Here, the minimum for criterion $F_{\text{ADD}}$ was arbitrarily set to $F_{\text{ADD}}(T_{\text{root}})$ at $T_{\text{root}} = 200$ because the profile decreased monotonically for all greater values of the root time. In all evaluations, we avoid putting arbitrary constraints near the root of the tree (for example, that the rate at the root is the same as that of one of its descendants) because while arbitrary constraints may bound the minimum, they will also create unknown biases. Note that the MVS-$F_{\text{IR}}$ profile is symmetric and parabolic around the estimated root time, whereas those of $F_{\text{LOG}}$ and $F_{\text{ADD}}$ are not. We see the same thing even in worked examples using a perfectly clock-like tree.

Our novel way of estimating CIs jointly takes into account some, but not all, of the sources of error noted in a previous publication [13]. These include errors in edge length estimation

(because of a finite sampling of sites), the unpredictability of fluctuating evolutionary rates, and coalescent time. The assumptions of this method, vis-à-vis independence of errors, are the same as those used in programs by Thorne et al. [31]. Note that even if edge lengths have no error due to sampling of sites (e.g., the type of error identified by the bootstrap), they may still show four wide CIs because of the other two factors, which do not go to zero with increasing sequence length. This, in turn, shows that using the sequence bootstrap to infer errors on divergence times [17,18] is not comprehensive and therefore the CIs are too narrow. We expect that this empirical approach to estimating standard errors will become increasingly accurate with increased numbers of internal nodes and calibration points.

Fortran source code and executable versions of the programs used in this study are downloadable from <http://www.kochi-ms.ac.jp/~ct_cmis/kitazoe/>. The operation of these programs, using the data in this analysis, is documented in Methods S2.

## SUPPORTING INFORMATION

**Methods S1** The procedure of MVS distance modification for placental mammals
Found at: doi:10.1371/journal.pone.0000384.s001 (0.04 MB PDF)

**Methods S2** The operation manual of programs for divergence time and MVS analyses
Found at: doi:10.1371/journal.pone.0000384.s002 (0.04 MB PDF)

**Figure S1** Trace of evolutionary rates along seven lineages in the MVS-$F_{IR}$ and ML-$F_{IR}$ analyses. The ML-$F_{IR}$ analysis (B) showed a flatter peak rate than did the MVS-$F_{IR}$ tree (A) (the black lines of F in the whale lineage) and inferred a longer period at a lower rate of evolution, producing an older root time. If lineages appear to merge with each other, zoom in to follow their exact path.
Found at: doi:10.1371/journal.pone.0000384.s003 (0.68 MB TIF)

**Table S1** Sequences and accession numbers used in this paper. The aligned amino acid sequences of the mitochondrial proteins ND1-ND6, ND4L, CO1-CO3, CTYB, and ATP6, are concatenated giving a total length of 3660 sites.
Found at: doi:10.1371/journal.pone.0000384.s004 (0.09 MB DOC)

**Table S2** Divergence times of the MVS-$F_{IR}$, ML-$F_{IR}$, and ML-$F_{LOG}$ analyses. The numbers 1–61 denote the ancestral nodes in Figure 1. The red numbers stand for the positions of the time constraints. Times older than the constraints varied the most, particularly using ML branch lengths.
Found at: doi:10.1371/journal.pone.0000384.s005 (0.09 MB DOC)

## Author Contributions

Analyzed the data: HK NN YK TO TW YO. Wrote the paper: HK PW YK. Other: Developed the initial theory and developed the simulations: HK YK. Developed the simulations, added biological/theoretical interpretations and revised the manuscript: PW. Developed the Fortran programs: NN TO TW YO.

## REFERENCES

1. Stanhope MJ, Waddell VG, Madsen O, de Jong W, Hedges SB, et al. (1998) Molecular evidence for multiple origins of Insectivora and a new order of endemic African insectivore mammals. Proc Natl Acad Sci U S A 95: 9967–9972.
2. Waddell PJ, Okada N, Hasegawa M (1999) Towards resolving the interordinal relationships of placental mammals. Syst Biol 48: 119–137.
3. Murphy WJ, Eizirik E, O'Brien SJ, Madsen, Scally OM, et al. (2001) Resolution of the early placental mammal radiation using Bayesian phylogenetics. Science 294: 2348–2351.
4. Waddell PJ, Kishino H, Ota R (2001) A phylogenetic foundation for comparative mammalian genomics. Genome Informatics 12: 141–154.
5. Springer MS, Stanhope MJ, Madsen O, de Jong WW (2004) Molecules consolidate the placental mammal tree. Trends in Ecology and Evolution 19: 430–438.
6. Easteal S (1999) Molecular evidence for the early divergence of placental mammals. BioEssays. 21: 1052–1058.
7. O'Brien SJ, Menotti-Raymond M, Murphy WJ, Nash WG, Wienberg J, et al. (1999) The promise of comparative genomics in mammals. Science 286: 458–462.
8. Arnason U, Adegoke JA, Bodin K, Born EW, Esa YB, et al. (2002) Mammalian mitogenomic relationships and the root of the eutherian tree. Proc Natl Acad Sci U S A 99: 8151–8156.
9. Benton MJ, Ayala FJ (2003) Dating the tree of life. Science 300: 1698–1700.
10. Alroy J (1999) The fossil record of North American mammals: evidence for a Paleocene evolutionary radiation. Syst Biol 46: 107–118.
11. Archibald JD, Deutschman DH (2001) Quantitative analysis of the timing of the origin and diversification of extant placental orders. J Mammal Evol 8: 107–124.
12. Kumar S, Hedges SB (1998) A molecular timescale for vertebrate evolution. Nature 392: 917–920.
13. Waddell PJ, Cao Y, Hasegawa M, Mindell DP (1999) Assessing the Cretaceous superordinal divergence times within birds and placental mammals by using whole mitochondrial protein sequences and an extended statistical framework. Syst Biol 48: 119–137.
14. Cao Y, Fujiwara M, Nikaido M, Okada N, Hasegawa M (2000) Interordinal relationships and timescale of eutherian evolution as inferred from mitochondrial genome data. Gene 23: 149–158.
15. Springer MS, Murphy WJ, Eizirik E, O'Brien SJ (2003) Placental mammal diversification and the Cretaceous–Tertiary boundary. Proc Natl Acad Sci U S A 100: 1056–1061.
16. Kimura M (1968) Evolutionary rate at the molecular level. Nature 217: 624–626.
17. Sanderson MJ (1997) Nonparametric approach to estimating divergence times in the absence of rate constancy. Mol Biol Evol 14: 1218–1231.
18. Sanderson MJ (2004) Estimating absolute rates of molecular evolution and divergence times: A penalized likelihood approach. Mol Biol Evol 19: 101–109, Software r8s, *version 1.71 homepage http://ginger.ucdavis.edu/*.
19. Horai S, Hayasaka K, Kondo R, Tsugane K, Takahata N (1995) Recent African origin of modern humans revealed by complete sequences of hominoid mitochondrial DNAs. Proc Natl Acad Sci U S A 92: 532–536.
20. Waddell PJ, Cao Y, Hauf J, Hasegawa M (1999) Using novel phylogenetic methods to evaluate mammalian mtDNA, including amino acid-invariant sites–LogDet plus site stripping, to detect internal conflicts in the data, with special reference to the positions of hedgehog, armadillo, and elephant. Syst Biol 48: 31–53.
21. Mouchaty SK, Gullberg A, Janke A, Arnason U (2000) The phylogenetic position of the Talpidae within eutheria based on analysis of complete mitochondrial sequences. Mol Biol Evol 17: 60–67.
22. Nikaido M, Cao Y, Harada M, Okada N, Hasegawa M (2003) Mitochondrial phylogeny of hedgehogs and monophyly of Eulipotyphla. Mol Phylogenet Evol 28: 276–284.
23. Zhang J, Kumar S (1997) Detection of convergent and parallel evolution at the amino acid sequence level. Mol Biol Evol 14: 527–536.
24. Swofford, DL, Olsen GJ, Waddell PJ, Hillis DM (1996) *Phylogenetic Inference.* In: Molecular Systematics, second edition, Hillis DM, Moritz, eds. Sunderland, Mass. Sinauer Assoc. pp 450–572.
25. Kitazoe Y, Kurihara Y, Narita Y, Okuhara Y, Tominaga A (2001) A new theory of phylogeny inference through construction of multidimensional vector space. Mol Biol Evol 18: 812–828.
26. Kitazoe Y, Kishino H, Okabayashi T, Watabe T, Nakajima N, et al. (2005) Multidimensional vector space representation for convergent evolution and molecular phylogeny. Mol Biol Evol 22: 704–715.
27. Okabayashi T, Kitazoe Y, Kishino H, Watabe T, Nakajima N, et al. (2006) Core set approach to reduce uncertainty of gene trees. BMC Evol Biol. 6: 41:10.1186/1471-2148-6-41.
28. Zuckerkandl E, Pauling L, Bryson V, Vogel HJ, eds (1965) Evolving genes and proteins. New York: Academic Press. 97 p.
29. Kishino H, Hasegawa M (1990) Converting distance to time: application to human evolution. Methods Enzymol. 183: 550–70.

30. Yoder AD, Yang Z (2000) Estimation of primate speciation dates using local molecular clocks. Mol Biol Evol 17: 1081–1090.

31. Thorne JL, Kishino H, Painter IS (1998) Estimating the rate of evolution of the rate of molecular evolution. Mol Biol Evol 15: 1647–1657.

32. Yang Z, Yoder A (2003) Comparison of likelihood and Bayesian methods for estimating divergence times using multiple gene loci and calibration points, with application to a radiation of cute looking mouse lemur species. Syst Biol 52: 705–716.

33. Hedges SB, Kumar S (2004) Precision of molecular time estimates. Trends in Genetics 20: 242–247.

34. Misawa K, Nei M (2003) Reanalysis of Murphy et al.'s data gives various mammalian phylogenies and suggests overcredibility of Bayesian trees. J Mol Evol 57: Supple 1: S290–S296.

35. Yang Z (1997) PAML: a program for phylogenetic analysis by maximum likelihood. Comput Appl Biosci 13: 555–556.

36. Jones DT, Taylor WR, Thornton JM (1992) The rapid generation of mutation data matrices from protein sequences. Comput Appl Biosci 8: 275–282.

37. Near TJ, Sanderson MJ (2004) Assessing the quality of molecular divergence time estimates by fossil calibrations and fossil-based model selection. Philos Trans R Soc Lond B Biol Sci 359: 1477–1483.

38. Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4: 406–425.

39. Drummond AJ, Ho SYW, Phillips MJ, Rambaut A (2006) Relaxed phylogenetis and dating with confidence. PLoS Biology 4: 699–710, Software BEAST *homepage http://evolve.zoo.ox.uk/beast/*.

40. Stein C (1956) Inadmissibility of the usual estimator for the mean of a multivariate distribution. Proc Third Berkeley Symp Math Statist Prob 1: 197–206.

41. Bromham L, Ramhaut A, Fortey R, Cooper A, Penny D (1998) Testing the Cambrian explosion hypothesis by using a molecular dating technique. Proc Natl Acad Sci U S A 95: 12386–12389.

42. Lynch M, Conery JS (2000) The evolutionary fate and consequences of duplicate genes. Science 290: 1151–1155.

43. Waters PD, Dobigny G, Waddell PJ, Robinson TJ (2007) Evolutionary History of LINE-1 in the Major Clades of Placental Mammals. PLoS ONE 2: (1): e158. doi:10.1371/journal.pone.0000158.